

# MDR을 이용한 XML DTD 이질성 해결 기법†

김진관<sup>o</sup> 김중일<sup>\*\*</sup> 최오훈<sup>\*</sup> 백두권<sup>\*</sup>

<sup>\*</sup>고려대학교 컴퓨터학과, <sup>\*\*</sup>라임미디어 테크놀러지스  
(jkkim<sup>o</sup>, pens, baik)<sup>o</sup>@software.korea.ac.kr, jikim@LimeMedia.co.kr,

## A Heterogeneity Resolution of XML DTD Using MDR

Jin-Kwan Kim<sup>o</sup>, John i Kim<sup>\*\*</sup>, O-Hoon Choi<sup>\*</sup>, Doo-Kwon Baik<sup>\*</sup>

<sup>\*</sup>Dept. of Computer Science & Engineering, Korea University

<sup>\*\*</sup>Lime Media Technologies Co., LTD.

### 요 약

정보통신기술의 급속한 발달과 인터넷 환경의 급속한 확산에 따라 정보통신 시스템의 통합과 다양한 분야의 정보통신 시스템 사이의 정보공유가 필수적이다. 그러나 정보통신 시스템의 통합과 정보공유의 가장 큰 걸림돌은 시스템의 통합 메커니즘 문제뿐만 아니라, 데이터의 의미(semantic), 구조(structure), 표현(representation)의 불일치이다. MDR은 정의된 표준 요소와의 매칭을 통해 데이터의 의미와 표현에 관한 이질성 문제를 해결할 수 있도록 해주며, XML은 메타데이터 레지스트리(Metadata Registry, MDR)에서 제공하지 않는 구조화된 데이터에 대한 표현 방법을 제공한다. 본 논문에서는 분야별 종적 데이터 공유를 위한 데이터의 의미, 구조, 표현의 이질성에 따른 문제를 분류하고, MDR과 XML의 활용을 통해 그 해결 방안을 제시한다.

### 1. 서론

정보통신기술의 급속한 발달과 인터넷 환경의 급속한 확산에 따라 새로운 비즈니스 환경에 적응하기 위해서는 기존의 정보통신 시스템의 통합과 다양한 분야의 정보통신 시스템 사이의 정보공유가 필수적이다. 그러나 정보통신 시스템의 통합과, 정보공유의 가장 큰 걸림돌은 시스템의 통합 메커니즘 문제뿐만 아니라, 데이터의 의미, 구조, 표현의 불일치이다.[3] 이에 대하여 데이터 이질성에 의한 충돌을 해결하고자 하는 분류 및 해결 방법이 연구되어 왔다.

현재 ebXML, BizTalk 등과 같이 시스템간의 수평적 통합 프레임워크(Framework)들이 출현하여, XML을 이용한 어플리케이션(Application)간의 통합이 대두되는 추세이다. XML은 메타데이터의 구조적인 표현을 가능하게 해준다. ebXML, BizTalk 등은 도메인이 다른 이질 데이터의 수평적 통합을 제공하는데, 비즈니스를 위한 데이터의 교환, 공유, 통합에는 수평적 통합뿐만 아니라 이질 데이터에 대한 수직적 통합도 중요하다.

ISO/IEC JTC1 WG2에서는 데이터 공유 및 교환을 위한 수직적인 해결방안으로 ISO/IEC 11179 - "메타데이터 레지스트리"를 통한 데이터의 의미, 표현의 표준화 방안을 제시하고 있다.[3] 본 논문에서는 MDR과 XML의 활용을 통해 분야별, 수직적 데이터 공유를 위한 데이터의 의미, 구조, 표현에 대한 이질성 문제를 분류하고 그 해결 방안을 제시한다.

### 2. 관련 연구

#### 2.1 충돌문제 분류와 스키마간 의미 연관

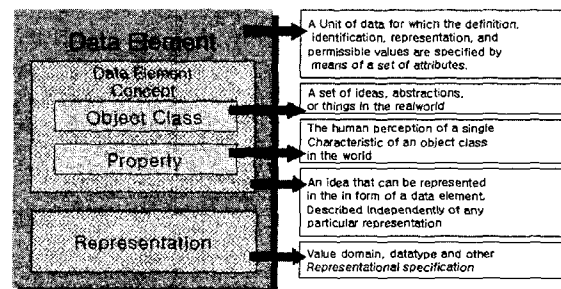
이질 데이터의 교환, 공유, 통합에 있어서 선행되어야 할 과제

† 본 논문은 정보통신부의 "지식 콘텐츠 기술을 위한 메타데이터 관리 시스템의 개발" 사업의 일부로 수행된 결과임

는 교환, 공유, 통합시에 발생하는 문제점에 대한 파악과 그 해결책 모색이다. 이와 관련하여 W.Kim과 J.Seo는 멀티 데이터베이스 시스템에서 데이터 충돌을 해결하기 위해 CDB(Component Database)간의 통합시에 일어날 수 있는 스키마와 데이터 이질성을 정의하고 이를 크게 Schema Conflicts와 Data Conflict의 두 가지로 구분하여, 테이블과 에트리뷰트, Data의 이질성으로 인한 문제를 분류하였다.[2] 또한 David wang은 XML로 데이터를 교환하는 다중의 이질 시스템간의 상호운용성 조정을 위해 스키마 요소간의 의미를 자동으로 연관시켜주는 X-MAP system에 대하여 제안하였다. X-MAP 시스템은 사람의 도움과 동등(equal)의 정의를 이용하여 스키마 요소간에 가능한 연관을 만들어 주는 반자동 시스템이다.[1]

#### 2.2 11179기반 MDR

정보통신 기술의 발달과 함께, 기존의 정보통신 시스템의 통합과, 정보공유에 있어서 시스템 레벨의 통합을 위한 메커니즘 개발뿐만 아니라, 데이터 자체에 대한 데이터 의미, 구분, 표현의 불일치 해결이 필요하다.



[그림 1] 데이터 요소의 구성

이러한 요소의 불일치 해결을 위해 ISO/IEC11179에 기반을 둔 MDR은 데이터 레벨에서의 통합을 위하여 데이터 요소와 그들 값의 의미, 표현, 구분을 기술한 정리된 디렉토리를 제공한다. 특히, MDR은 명세에 맞는 데이터만을 가공하여 등록하고 관리하므로, 메타데이터 집합에서 기술된 데이터에 대하여 공통된 이해를 유지하고 의미의 중복성을 방지한다. 또한 MDR은 데이터 공유를 위한 기본틀을 제공한다. [그림1][3]

3. 본 론

3.1 데이터의 교환, 공유, 통합을 위한 정보

데이터의 교환, 공유, 통합을 위해서는 데이터에 대한 정보가 필요하다. 데이터에 대한 정확한 정보는 데이터가 가지고 있는 의미, 표현, 그리고 메타데이터가 가지고 있는 구조를 파악함으로써 얻을 수 있다.[3] 메타데이터가 가지고 있는 구조 정보가 필요한 이유는 정보를 교환, 공유, 통합하고자 하는 각각의 로컬 데이터베이스 상의 테이블구조와 구성이 모두 상이하기 때문이다.

XML 문서로 데이터를 전달할 경우, 로컬 데이터베이스의 테이블에서 표현하고 있는 스키마의 속성들의 구조대로 XML 문서를 작성하게 된다면, 다른 로컬 데이터베이스에서는 그 구조를 이해할 수 없으므로 올바른 데이터를 추출해 낼 수가 없다. 따라서 3가지 관점에서 데이터의 이질성을 파악하여야 한다.

Semantic	데이터가 가지고 있는 요소의 의미 정보
Schematic	메타데이터가 가지고 있는 구조 정보
Representation	데이터의 값 영역, 데이터 형, 측정 단위 등의 표현 정보

[표 1] 데이터 표현의 3요소

3.2 데이터 이질성 항목 분류

데이터의 교환, 공유, 통합시에 발생하는 문제점들을 해결하기 위해서는 데이터 이질성에 대한 분류가 선행되어야 한다. [표 2]는 XML 문서로 데이터를 전달할 때 MDR을 이용하여 등록된 표준 요소들과 로컬 데이터베이스의 데이터 정보 간에 일어날 수 있는 데이터 이질성을 [표 1]의 데이터 표현 3요소에 근거하여 분류하였다.

Data Heterogeneity		Definition
Semantic	Substitutable*	하나의 엘리먼트가 다른 엘리먼트로 대체된다.
	Composed	하나의 엘리먼트가 다른 엘리먼트들의 집합으로 구성된다.
Schematic	Composing	하나의 엘리먼트가 다른 엘리먼트를 구성하는 요소들 중 하나이다.
	Rearranging	하나의 엘리먼트가 다른 엘리먼트와 구조를 이루는 순서가 다르다.
Representation	MeasureUnit*	하나의 엘리먼트가 다른 엘리먼트와 측정 단위가 다르다.
	Enumeration*	하나의 엘리먼트가 다른 엘리먼트와 다른 코드셋을 가진다.

\* 표시는 MDR만으로 해결할 수 있는 데이터 이질성 항목을 나타냄

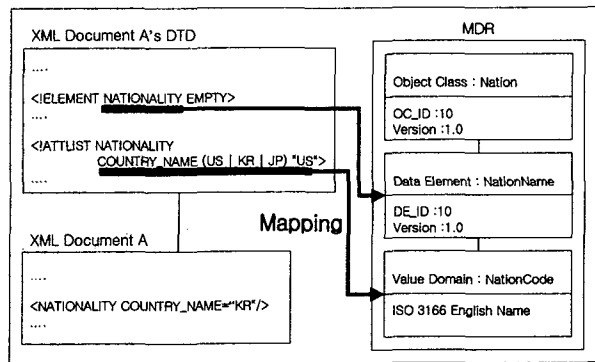
[표 2] 데이터 이질성 분류

3.3 데이터 이질성 문제 해결 방법

Semantic과 Representation에 관한 이질성 문제는 MDR에 정의된 표준 요소와의 매칭을 통하여 해결할 수 있다. MDR에 의해 관리되는 데이터의 기본단위는 데이터 요소이다. 데이터 요소는 3가지 구성요소로 이루어진다. 각각의 구성요소는 [그림 1]의 object class, property, representation이다. 이 3가지 구성요소를 통하여 각 데이터에 대한 규격화된 표현방식, 각 데이터를 구분할 수 있는 속성, 각 데이터가 속하는 도메인에 대한 표준 정의를 한다. 이러한 표준화된 데이터 요소를 통하여 동일한 semantic을 갖는 데이터간의 substitutable 충돌 문제를 해결한다. 또한 representation의 MeasureUnit, Enumeration 충돌 문제는 데이터 요소간 값 사상 기법을 통해 해결된다.[5]

3.3.1 11179기반 MDR을 이용한 의미, 표현 정보의 전달

[그림 2]은 XML DTD로 표현된 로컬 스키마를 MDR과 연관 지어주는 모델이다. DTD의 요소는 MDR의 Data Element와 매핑되고, 데이터의 타입은 Value Domain과 매핑되어 의미와 표현 방식을 파악한다. 그러나 XML DTD에서 표현하는 데이터의 구조는 표현할 수 없으므로 구조표현에 대한 문제점은 다른 방식으로 해결해야만 한다.



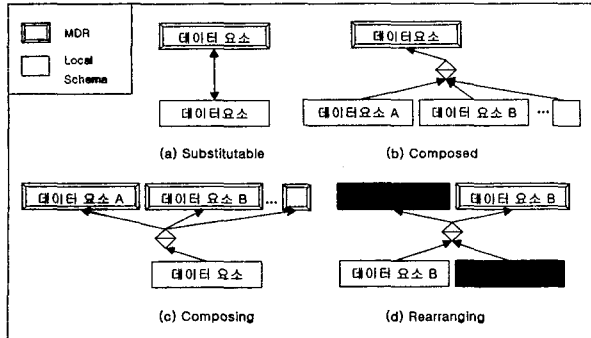
[그림 2] DTD와 MDR간의 의미연관

3.3.2 XML을 이용한 메타데이터 구조정보 전달

MDR만으로는 데이터의 정보 중 한 요소인 메타데이터가 가지고 있는 구조 정보를 나타낼 수 없다. 따라서 MDR과 함께 이용할 수 있는 다른 해결방안이 필요하다. XML은 엘리먼트와 엔터티, 그리고 애트리뷰트를 이용하여 데이터뿐만 아니라 데이터와 데이터가 가지고 있는 구조와 형식을 정의할 수 있다. 따라서 XML을 이용하여 데이터베이스간의 정보를 전달할 경우 로컬 데이터베이스에서 가지고 있는 메타데이터의 구조 정보를 함께 표현할 수 있다.

예를 들어, 로컬 데이터베이스 A와 B가 있다고 가정할 때, 테이블에 사람의 이름을 나타내는 필드가 있다고 하자. 이 때, A는 Name이라는 하나의 애트리뷰트로 나타내고, B는 FirstName과 LastName으로 나누어 관리하는 경우가 있을 수 있다. 이런 경우 B의 두 애트리뷰트를 합쳐야만 A의 Name을 표현할 수 있다. 따라서 A와 B간의 데이터 전달을 위해서는 메타데이터에 대한 구조정보가 꼭 필요하다. 메타데이터의 구조 정보는 XML Schema, DTD, RDF 등을 통해서 표현할 수 있다. 본 논문에서는 가장 접근하기 쉬운 DTD를 이용하여 해결방안을 제시한다.

로컬 데이터베이스의 스키마 속성은 XML의 데이터 요소로 바꾸어 전달한다. 가장 기본적인 형태의 구조는 데이터 요소간 1:1 매핑이 가능한 형태 - “[그림 3]의 (a)”이다.



[그림 2] 데이터 요소의 구조적 불일치 분류

3.3.2.1 Composed

하나의 요소가 다른 요소들의 집합으로 이루어진 경우 - “[그림 3]의 (b)”이다. XML 문서로 데이터를 전달하고자 할 때, D 관련된 여러 개의 로컬 스키마 속성들을 DTD의 요소로 정의하고, 그것들을 하나로 묶어 상위의 요소를 생성한 뒤, 그 상위 요소를 매칭시키고자 하는 데이터 요소와 연관지어 이질성 문제를 해결한다.

[Function 1]~[Function 3]은 KIF(Knowledge Interchange Format)으로 표현한 mapping rule이다.[4]

[Function 1]을 표현하기 위한 명제는 다음과 같이 정의된다.

```

same(x, y) <= x = y
get_element(S, a) <= extract a from S
set_element(S(x), a) <= link a to S(x)
z = +(x, y)
    
```

[Function 정의용 위한 명제]

```

set_element(MDR(z),+(get_element(DTD, x), get_element(DTD, y)))
<= same(get_element(MDR,z),
+(get_element(DTD,x),get_element(DTD,y)))
    
```

[Function 1] Composed Structure Mapping Rule

3.3.2.2 Composing

하나의 요소가 다른 요소를 구성하는 요소들 중 하나인 경우 - “[그림 3]의 (c)”이다. DTD를 작성할 때, MDR에서 제공하는 요소에 부합하는 여러 개의 하위 요소를 만든 뒤, DTD와 MDR에서 같은 의미를 갖는 각각의 요소를 연관지어 문제를 해결한다. 이 경우는 다른 경우와 달리 자동화할 수 없다. 따라서 로컬 데이터베이스 스키마 속성을 DTD에 나누어 기록하는 부분에서 사람의 도움이 추가되는 semi-automated 시스템이 가능하다.[4]

```

set_element(MDR(x'),x) and set_element(MDR(y'),y)
<= same(get_element(DTD,z),
+(get_element(MDR,x'),get_element(MDR,y')))
    
```

[Function 2] Composing Structure Mapping Rule

3.3.2.3 Rearranging

데이터 요소간의 순서가 존재하는 경우 - “[그림 3]의 (d)”이다. 이 경우는 데이터 요소가 여러 개의 로컬 스키마의 속성으로 구성되는 경우이다. Rearranging은 Composed나 Composing의 경우와 함께 발생한다. 이것은 [Function 1]이나 [Function 2]를 수행한 후에 [Function 3]을 이용하여 하위 요소의 순서를 재배치하여 문제를 해결한다.[4]

```

set_element(MDR(x),y') and set_element(MDR(y),x')
<= same(get_element(MDR,x), get_element(DTD,y'))
and same(get_element(MDR,y), get_element(DTD,x'))
    
```

[Function 3] Rearranging Structure Mapping Rule

3.3.2.4 Compound

이러한 구조의 이질성 문제는 단독으로 나타나기도 하지만, 대부분의 경우 두 가지 이상이 복합적으로 나타나거나, 중첩되기도 한다.[2] 복합적으로 나타나는 경우에는 [Function 1]부터 [Function 3]까지를 하나씩 차례대로 적용하여 해결할 수 있다. 구조가 [그림 3]과 같이 단순한 2계층이 아닌, 여러 층으로 이루어졌을 경우는 [Function 1]과 [Function 2]를 여러 번 반복하여 적용함으로써 더 복잡한 구조를 나타낼 수 있다.

4. 결론 및 향후 연구 과제

데이터의 교환, 공유, 통합을 위해서는 데이터에 대한 정보가 필요하다. 데이터에 대한 정확한 정보는 데이터가 가지고 있는 의미, 표현, 그리고 메타데이터가 가지고 있는 구조를 파악해야 한다. MDR은 데이터에 대한 의미와 표현에 대한 표준을 제공함으로써 기존 방식의 데이터 통합이 가지고 있던 문제점들을 보다 현실적으로 해결해 주지만, 메타데이터에 대한 구조정보 포함하고 있지 않기 때문에 MDR만으로는 구조적 불일치 문제를 해결할 수 없다. 따라서 본 논문에서는 데이터 교환, 공유, 통합시에 발생하는 충돌 문제에 대한 분류와 함께 MDR과 XML을 이용한 해결책을 제시하였다. XML은 ebXML과 같은 수평적 통합 방법론과 MDR을 이용한 수직적 표준간의 조화를 가능하게 한다. 이를 위해 XML을 이용한 충돌 문제의 해결 방법 구현에 대한 연구를 진행하고 있다.

5. 참고문헌

[1] David Wang, Automated Semantic Correlation between Multiple Schema for Information Exchange, M.I.T., MM, May 2000  
 [2] W..Kim and J.Seo, Classifying schematic and data heterogeneity in multi-database systems. IEEE Computer 24(12):12-18, 1991  
 [3] ISO/IEC IS 11179, Information technology - Specification and standardization of data elements  
 [4] http://logic.stanford.edu/kif/kif.html  
 [5] ISO 11179기반 데이터 레지스트리에서 데이터 요소간 값 사상, 김승훈, 박대하, 나홍석, 백두권, 한국정보과학회 춘계 학술발표논문집(B권), pp.48~50, 1999